

Distribution of Four Founding mtDNA Haplogroups Among Native North Americans

JOSEPH G. LORENZ AND DAVID GLENN SMITH

Department of Anthropology, University of California, Davis, California 95616

KEY WORDS PCR, Migration, Language group

ABSTRACT The mtDNA of most Native Americans has been shown to cluster into four lineages, or haplogroups. This study provides data on the haplogroup affiliation of nearly 500 Native North Americans including members of many tribal groups not previously studied. Phenetic cluster analysis shows a fundamental difference among 1) Eskimos and northern Na-Dene groups, which are almost exclusively mtDNA haplogroup A, 2) tribes of the Southwest and adjacent regions, predominantly Hokan and Uto-Aztecan speakers, which lack haplogroup A but exhibit high frequencies of haplogroup B, 3) tribes of the Southwest and Mexico lacking only haplogroup D, and 4) a geographically heterogeneous group of tribes which exhibit varying frequencies of all four haplogroups. There is some correspondence between language group affiliations and the frequencies of the mtDNA haplogroups in certain tribes, while geographic proximity appears responsible for the genetic similarity among other tribes. Other instances of similarity among tribes suggest hypotheses for testing with more detailed studies. This study also provides a context for understanding the relationships between ancient and modern populations of Native Americans. © 1996 Wiley-Liss, Inc.

The mtDNA of most Native Americans has been shown to cluster into four haplogroups, designated A, B, C, and D, that define four separate mtDNA lineages (Schurr et al., 1990; Torroni et al., 1992). Other haplogroups have been identified that might have contributed to the founding population of Native Americans (e.g., haplogroup E [Bailliet et al., 1994]), but there is evidence that at least some of these are due to recent mutations or are of non-Native American origin (Torroni and Wallace, 1995) and in any case represent a minority of Native Americans (i.e., 4.1% [Bailliet et al., 1994]). The haplogroups that define these lineages are identified by the presence or absence of characteristic restriction enzyme cleavage sites (i.e., the presence of a *Hae* III site at nucleotide pair [np] 663 defines haplogroup A, the absence of a *Hinc* II site at np 13,259 and the presence of an *Alu* I site at np 13,262 defines

haplogroup C, and the absence of an *Alu* I site at np 5,176 defines haplogroup D) or, in the case of haplogroup B, the presence of a 9 bp deletion in the intergenic region between the cytochrome oxidase II and lysine tRNA genes (Schurr et al., 1990; Torroni et al., 1992, 1993a).

Haplogroups A, B, C, and D are also present in east Asia, where they have been identified among Koreans, Malaysian Chinese, Taiwanese Han (Ballinger et al., 1992) and Tibetans (Torroni et al., 1994a), although at much lower frequencies than in New World

Received March 1, 1995; accepted June 3, 1996.

Address reprint requests to David Glenn Smith, Department of Anthropology, University of California, Davis, California 95616.

Joseph G. Lorenz's current address is Analytical Genetic Testing Center, Suite 201, 7808 Cherry Creek S. Dr., Denver, CO 80231.

populations, leading to speculation of a founder effect on the New World gene pool (Wallace et al., 1985). Haplogroups A, C, and D are absent among non-Chinese southeast Asians. In Siberia haplogroup B is absent, but the frequencies of haplogroups A and especially C and D are higher than in east Asia (Torroni et al., 1993b). Torroni et al. (1992, 1993a) have shown that in North America the distribution of the haplogroups in northern Na-Dene tribal groups differs from that in Amerinds and southern Na-Dene tribal groups (the terms *Na-Dene* and *Amerind* refer to two of the three major linguistic "families" from Greenberg's [1987] taxonomy of Native American languages). Virtually all members of northern Na-Dene tribes belong to haplogroup A, while all four haplogroups have been reported to be present in the Amerinds and southern Na-Dene. However, the frequency of the four haplogroups varies greatly among Amerind ethnic groups, with some haplogroups being absent in some populations (Torroni et al., 1993a). In support of Greenberg (1987), it has been proposed that at least some haplogroups represent separate founding populations and that tribalization resulted from the subsequent isolation among incipient ethnic groups (Wallace et al., 1985; Torroni et al., 1993a). Hence, groups that speak similar languages should also exhibit genetic similarities (Torroni et al., 1994b; Torroni and Wallace, 1994). The conclusion of Chakraborty and Weiss (1991) that mtDNA of the New World is in mutation-drift equilibrium due to migration of lineages across population boundaries, while based on a very limited sample, is not consistent with the hypothesis of founder effects from multiple migrations. However, because studies of mtDNA variation in North America have focused mainly on Native American groups from the Northwest Coast and the American Southwest, mtDNA variation in tribes from a sufficient geographic range in North America to evaluate the fit of language and genetic similarities has not been characterized.

In an earlier study (Lorenz and Smith, 1994) we showed that the 9 bp deletion, a marker for haplogroup B, is present among all Amerind groups but that its frequency

varies considerably among these groups. Moreover, we found that geographic proximity influences the frequency of this haplogroup more strongly than does linguistic affiliation, probably due both to substantial female migration among adjacent but linguistically unrelated (or only remotely related) tribes and to language spreads across regions containing genetically unrelated (or only remotely related) tribes (Nichols, 1992). The purpose of the present study is not to address the issue of how the New World was settled but rather to extend our knowledge of the distribution of mtDNA variation among Native North Americans by determining whether or not the distribution of haplogroups A, C, and D, like that of haplogroup B, are also structured linguistically or geographically. This knowledge will be useful for reconstructing mtDNA relationships among living Native North Americans and provide a context within which to interpret the distribution of mtDNA haplogroups from archaeological populations (e.g., see Stone and Stoneking, 1993; Parr et al., 1995; Kaestle, 1995).

MATERIAL AND METHODS

Haplogroup determination

The haplogroup affiliation of 497 individuals from more than 40 ethnic groups across North America was determined. The source and description of these samples have been presented elsewhere (Lorenz and Smith, 1994). Most of the samples had previously been typed for the presence of the 9 bp deletion that defines haplogroup B (Lorenz and Smith, 1994), but an additional 101 samples were subsequently typed for the 9 bp deletion in order to increase the sample size of underrepresented tribal groups. All samples which were determined not to be haplogroup B (i.e., those that were determined to lack the 9 bp deletion) were amplified and tested for membership in haplogroup A. Those samples identified as belonging to haplogroup A (but not haplogroup B) were not amplified and tested for the polymorphic markers characterizing haplogroups C and D. Similarly, samples not assigned to haplogroup A that possessed the restriction site characterizing haplogroup C were not scored for haplo-

group D. Samples that did not exhibit any of the polymorphic markers characteristic of one of the four Native North American haplogroups were scored as *other*.

Because all 497 individuals were not typed for all four haplogroup markers, individuals with markers that are characteristic of two (or more) haplogroups would have been overlooked. Since the haplogroup affiliation of such "compound haplotypes" is ambiguous without more detailed analysis, they should be eliminated from comparisons of the distributions of haplogroups. However, compound haplotypes are relatively rare. Of the 1,102 individuals typed in Siberia and the New World in previous studies, only 11 had compound haplotypes (i.e., less than 1% [Schurr et al., 1990; Torroni et al., 1993a,b, 1994b; Bailliet et al., 1994]). It is unlikely that our failure to identify and eliminate the fewer than a half dozen compound haplotypes we would have expected to identify could distort the overall picture this study is meant to provide of the distribution of the four haplogroups in North America. Nevertheless, to ensure that compound haplotypes were not unexpectedly common in our own sample, 52 individuals included in the present study were typed for all four haplogroups.

The primers used to amplify regions of the mtDNA which contain the diagnostic restriction enzyme digestion sites for haplogroups A, B, C, and D are listed in Table 1. Amplification reactions were carried out in a total volume of 25 μ l and contained 0.5 μ l of each primer (1.0 μ M final concentration), 0.65 μ l of 20 mg/ml bovine serum albumin (Boehringer-Mannheim, Indianapolis, IN), and 12.5 μ l of PCR Master (Boehringer-Mannheim; 2.5 U Taq DNA polymerase in Brij 35, 0.005% [v/v], dATP, dCTP, dGTP, dTTP each at 0.2 mM, 10 mM Tris-HCl, 50 mM KCl, 1.5 mM MgCl₂ final concentration). The thermal cycling profile consisted of an initial denaturing for 3 min at 94°C followed by 35 cycles of 1 min at 94°C, 1 min at 45°C, and 1 min at 72°C. A 5 min 72°C extension step followed the last cycle. A 5 μ l portion of the amplification product was electrophoresed on a 6% polyacrylamide gel and stained with ethidium bromide to confirm the presence of PCR product. In cases where no visible product was detected, an internal amplification was

performed using 1 μ l of the previously amplified product in a reaction mixture identical to that described above except for the use of internal primers (see Table 1). The thermal cycling profile for the internal amplifications was as described above except that the primer annealing temperature was lowered to 37°C. In cases of successful amplification the remainder of the amplification reactions was digested for 1–3 h with 5–10 units of the appropriate restriction enzyme. A 5–10 μ l portion of the digest was then electrophoresed on a 6% polyacrylamide gel, stained with ethidium bromide, and photographed over a UV transilluminator to determine whether or not the amplified fragment had been cut by the restriction enzyme.

Individuals assigned a haplogroup affiliation based on the presence or absence of a restriction site in one of the amplified segments were grouped by tribe as illustrated in Table 2. All analyses described below were based on the data obtained from this study, our earlier study (Lorenz and Smith, 1994), and 332 individuals whose haplogroups were reported in other studies cited in Table 2.

Statistical analyses

Between-group heterogeneity. Each haplogroup was regarded as one of four alleles at a single, haploid locus. Individuals belonging to other haplogroups were removed from their respective tribes, and the frequencies of the four haplogroups in those tribes were rescaled to unity. Values of $\hat{\theta}$, which is equivalent to Wright's F_{ST} statistic adjusted for sampling bias (Weir, 1990), were then calculated to assess among-cluster variation in the frequencies of the four haplogroups after grouping the tribes in various configurations. $\hat{\theta}$ values for different configurations of clusters were calculated only to indicate whether clustering tribes in a given configuration increased or decreased, relative to a standard value, the heterogeneity among the clusters. Higher values of $\hat{\theta}$ indicate greater intragroup homogeneity and greater intergroup heterogeneity which is expected when the clustering of samples into groups is based on valid assumptions. Lower values of $\hat{\theta}$, on the other hand, indicate less intragroup homogeneity coupled with less intergroup heterogeneity, suggesting that the

TABLE 1. PCR primers used to amplify the four mtDNA haplogroup markers

Haplogroup	Polymorphic site	Primer coordinates ¹	Primer sequence (5'→3')
A	<i>Hae</i> III-663	535-553	CCCATACCCCGAACCAACC
		725-707	GGTGAACCTCACTGGAACGG
		568-577	CCCCCCACAG
		701-692	CTTGCAATGTG
B	9 bp deletion	8,150-8,166	CCGGGGGTATACTACGG
		8,366-8,345	TTTCACTGTAAAGAGGTGTTGG
		8,316-8297	ATGCTAAGTTAGCTTTACAG
		8,196-8,215	ACAGTTTCATGCCATCGTC
C	<i>Hind</i> II-13259	13,197-13,213	GCAGCAGTCTGCGCCC
	<i>Alu</i> I-13262	13,403-13,384	ATATCTTGTTTCATTGTAA
		13,236-13,245	AATCGTAGCC
		13,374-13,366	GATGGACCC
D	<i>Alu</i> I-5176	5,151-5,170	CTACTACTATCTCGCACCTG
		5,481-5,464	GTAGGAGTAGCGTGGTAA
		5,161-5,170	CTCGCACC
		5,272-5,263	TCGATAATGG

¹ Numbered according to the Anderson et al. (1981) reference sequence.

assumptions underlying that grouping lack validity. If configurations of clusters that conform to language taxonomy but crosscut many geographic regions are found to have high $\hat{\theta}$ values, a high degree of overlap between genetics and culture is indicated. This outcome is consistent with a pattern of early biologic differentiation which was paralleled by linguistic differentiation and then followed by isolation with only very recent, if any, admixture. High $\hat{\theta}$ values for geographic configurations that crosscut language affiliations, on the other hand, provide evidence for gene flow (and/or recent genetic differentiation) within a region. Low values of $\hat{\theta}$ for configurations of clusters conforming to both language taxonomy and geographical proximity are consistent with language spreads not associated with admixture. High values of $\hat{\theta}$ for configurations of clusters conforming to both language taxonomy and geographical proximity are consistent with early tribalization followed by recent admixture.

As a standard for comparing values of $\hat{\theta}$ associated with alternative configurations of clustering, a baseline $\hat{\theta}$ value was calculated for 30 tribes. Some of these 30 unclustered tribes actually comprise pools of tribes included in Table 2 that are both geographically and linguistically proximate but for which few samples were available. For example, Chippewa and Kickapoo were pooled together since they both speak Algonquian languages and they are approximately geographically contiguous (Chippewa individu-

als were sampled from South Dakota, Minnesota, and Wisconsin; the Kickapoo homeland is in Wisconsin). Figure 1 indicates the geographic location of the 30 tribes. Tribes represented by fewer than 10 individuals and which could not be pooled using the criteria described above were not included in this baseline analysis (e.g., the Micmac individuals, although representing a tribe that speaks an Algonquian language, were not pooled with the Chippewa/Kickapoo because they live in Newfoundland).

For geographic analysis all the tribes were divided into one of the following areas as defined in our previous study (Lorenz and Smith, 1994): Arctic/Subarctic, Northwest Coast, East, Midwest/Great Plains, Southwest, California/Great Basin, and Mexico. Four identical analyses were conducted on hierarchically arranged clusters of tribes that were variously grouped (i.e., 3, 9, 20, and 24 language groups) according to language affiliation (Campbell and Mithun, 1979; Greenberg, 1987). The statistical significance of differences in the frequency distributions of haplogroups among clusters of tribes based on geographic or language criteria was confirmed using the chi-square test for homogeneity (Siegel, 1956).

Genetic distance. The method of Cavalli-Sforza and Edwards (1967) was used to calculate genetic distances between pairs of the 30 modern tribes or tribes that had been pooled as described above and in addition

TABLE 2. Frequency of mtDNA haplogroups in modern Native North Americans

Linguistic affiliation			Frequency of haplogroups (%)							
Greenberg (1987)		Campbell and Mithun (1979)	Tribe	Sample size	A	B	C	D	Other	Reference ¹
Family	Sublevel 1									
Eskimo-Aleut Na-Dene	Inuit	Inuit	Inuit	30	0.967	0	0	0.033	0	a
	Haida	Haida	Haida	29	0.966	0	0	0.034	0	a,c (25)
	Tlingit	Tlingit	Tlingit	2	1	0	0	0	0	b
	Athabaskan-Eyak	Athabaskan	Dogrib	42	1	0	0	0	0	a,b (30)
Amerind	Athabaskan-Eyak	Athabaskan	Hupa	2	0	1	0	0	0	a
	Athabaskan-Eyak	Athabaskan	Navajo	58	0.517	0.414	0.034	0	0.034	a,b (48)
	Athabaskan-Eyak	Athabaskan	Apache	29	0.621	0.172	0.138	0.069	0	a,c (25)
	Almosan-Keresiouan	Algonquian	Chippewa/Kickapoo	62	0.484	0.113	0.194	0	0.21	a,c (43)
	Almosan-Keresiouan	Algonquian	Cheyenne/Arapahoe	26	0.308	0.115	0.346	0.154	0.077	a
	Almosan-Keresiouan	Algonquian	Micmac/Narragansett	7	0.286	0	0.143	0	0.571	a
	Almosan-Keresiouan	Yurok	Yurok	1	0	1	0	0	0	a
	Almosan-Keresiouan	Salishan	Bella Colla	36	0.5	0.056	0.139	0.25	0.056	a,c (25)
	Almosan-Keresiouan	Wakashan	Nuu-Chah-Nulth	15	0.4	0.067	0.133	0.267	0.133	c
	Almosan-Keresiouan	Caddoan	Pawnee	3	0.333	0.667	0	0	0	a
	Almosan-Keresiouan	Iroquois	Cherokee	16	0	0.313	0.313	0	0.375	a
	Almosan-Keresiouan	Keres	Acoma	1	1	0	0	0	0	a
	Almosan-Keresiouan	Siouan	Siouan	34	0.529	0.176	0.147	0.059	0.088	a
	Penutian	Penutian	California Penutian	17	0.118	0.412	0.059	0.412	0	a
	Penutian	Maya	Maya	27	0.519	0.222	0.148	0.074	0.037	b
	Penutian	Mixe-Zoquean	Mixe	16	0.625	0.313	0.063	0	0	d
	Penutian	Coos	Coos	1	0	0	1	0	0	a
	Penutian	Muskogean	Chickasaw/Choctaw	27	0.667	0.222	0.074	0	0.037	a
	Penutian	Muskogean	Creek/Seminole	18	0.389	0.111	0.167	0.167	0.167	a
	Penutian	Hokan	Zuni	22	0.182	0.636	0.091	0	0.091	a
Hokan	Hokan	Northern California Hokan	5	0	0.4	0.2	0.4	0	a,b (1)	
Hokan	Hokan	Washo	28	0	0.536	0.357	0.107	0	a	
Hokan	Yuman	Quechan/Cocopa	23	0	0.652	0.304	0	0.043	a	
Hokan	Yuman	Havasupai/Hualapai/Yavapai/Mojave	18	0.111	0.50	0.389	0	0	a	
Hokan	Yuman	Kumiai	16	0	0.625	0.375	0	0	a	
Hokan	Yuman	Kiliwa/Paipai	11	0	0.909	0.091	0	0	a	
Hokan	Yuman	Cochimi	13	0.077	0.462	0.462	0	0	a	
Hokan	Salinan/Chumash	Salinan/Chumash	11	0.455	0.182	0.091	0.273	0	a	
Kiowa-Tanoan	Kiowa-Tanoan	Kiowa	5	0.4	0	0.2	0	0.4	a	
Kiowa-Tanoan	Kiowa-Tanoan	Jemez/Taos/San Idelfonso	36	0	0.861	0.028	0.028	0.083	a	
Oto-Mangue	Mixtec	Mixtec	29	0.828	0.103	0.069	0	0	d	
Oto-Mangue	Zapotec	Zapotec	15	0.333	0.333	0.333	0	0	d	
Uto-Aztecan	Uto-Aztecan	Nahua/Cora	32	0.531	0.344	0.063	0	0.063	a	
Uto-Aztecan	Uto-Aztecan	Pima	37	0.054	0.568	0.378	0	0	a,b (30)	
Uto-Aztecan	Uto-Aztecan	Paute/Shoshone	9	0	0.222	0.222	0.444	0.111	a	
Uto-Aztecan	Uto-Aztecan	Hopi	4	0	1	0	0	0	a,b (1)	
Uto-Aztecan	Uto-Aztecan	Comanche	2	0	1	0	0	0	a	
Uto-Aztecan	Uto-Aztecan	California Uto-Aztecan	14	0	0.286	0.429	0.286	0	a	
			829	0.417	0.302	0.158	0.063	0.06		
			779	0.444	0.321	0.168	0.067			

¹References are as follows: a, present study; b, Torroni et al., 1992; c, Torroni et al., 1993a; d, Torroni et al., 1994b. In cases where the sample for an ethnic group is derived from more than one study, the number in parentheses denotes the sample size reported in the earlier study.

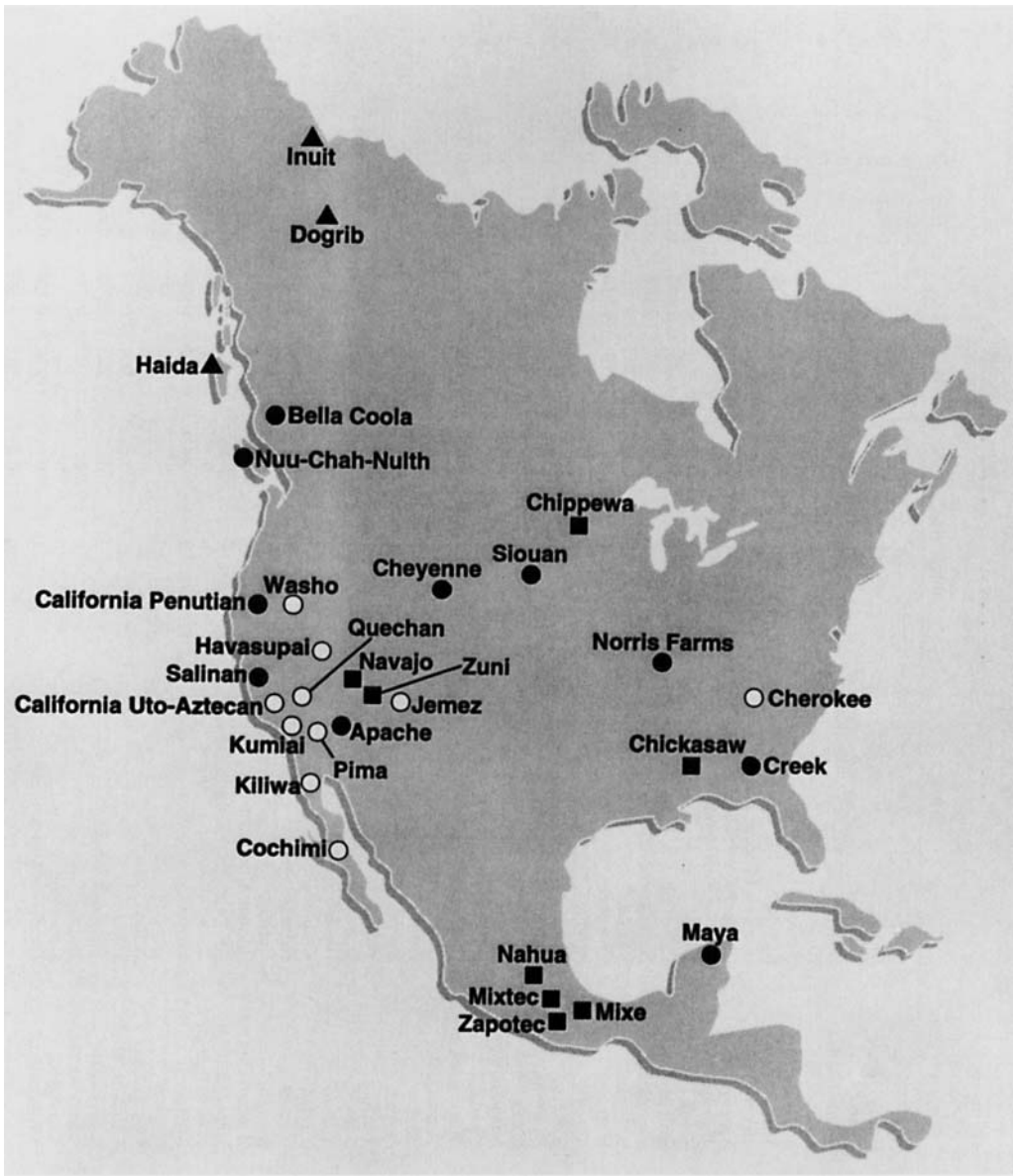


Fig. 1. Geographic location of the 30 modern Native American populations and one pre-Columbian population (Norris Farms [Stone and Stoneking, 1993]) used in the phenetic cluster analysis. Populations in phenetic cluster I are denoted by triangles, phenetic cluster II by squares, phenetic cluster III by filled circles, and phenetic cluster IV by open circles.

the archaeological population from Norris Farms in central Illinois which dates from AD 1300 (Stone and Stoneking, 1993). These distances were calculated from the frequencies of the four mtDNA haplogroups in each tribe after removing individuals belonging

to other haplogroups and rescaling the frequencies in those tribes to unity. To determine patterns of similarity among North American groups in the relative frequencies of the four haplogroups independent of language and geography, UPGMA and maxi-

mum likelihood trees were constructed using the CONTCHAR and DISTANCE programs of PHYLIP (Felsenstein, 1993).

RESULTS

mtDNA haplogroup distribution

The total number of Native North Americans typed for one of the four haplogroups or other is 497 (Lorenz and Smith, 1994; present study). None of the 52 samples that were typed for all four haplogroups exhibited a compound haplotype (i.e., belonged to more than a single haplogroup or other). The distribution of the frequencies of the four haplogroups and others identified in this study in addition to data on 332 Native American individuals from earlier studies is shown in Table 2. Haplogroup A was the most common (over 40%), and haplogroup D was the least common (less than 7%) of the four haplogroups. Fifty individuals, fewer than for any haplogroup, were classified as other, some of which probably result from admixture with other, presumably European, populations. That more than two-thirds of the others were from Eastern and Midwestern tribes, with whom earlier and more intense contact with nonnative, especially European, peoples occurred suggests that many of these others represent nonnative matrilineal (i.e., admixture with nonnatives). However, Stone and Stoneking (1993) and Parr et al. (1996) classified at least one individual as other (referred to as haplogroup N) in the studies of pre-Columbian populations in Illinois and Utah, respectively. Thus, assuming a recent mutation is not responsible, native haplogroups other than A, B, C, and D, while rare, apparently existed. A higher resolution screening of these other haplogroups might indeed show some to be Native American in origin, probably resulting from recent mutation at the diagnostic haplogroup restriction sites. However, haplotypes identified by high resolution analysis (e.g., see Torroni et al., 1993a) tend to be autapomorphies and hence of no use in determining relatedness among groups, and they were excluded from the analysis whose results follow.

Statistical analyses

Between-group heterogeneity. The overall $\hat{\theta}$ value, indicating the relative levels of

intergroup heterogeneity for the 30 unclustered tribes, is 0.291. As illustrated in Table 3, the relative intergroup heterogeneity varies greatly depending on how the tribes are clustered for estimating $\hat{\theta}$. When the tribes are partitioned into the seven geographic regions illustrated in Figure 2, $\hat{\theta}$ decreases from 0.291 to 0.193, indicating that the intraregion genetic homogeneity is substantially less than the intratribe genetic homogeneity, but the distribution of the four haplogroups is still heterogeneous among the seven geographic regions ($\chi^2 = 301.7$, 18 d.f., $P < 0.001$). However, when the entire Arctic/Subarctic region, in which haplogroup A is virtually fixed, is compared to a single cluster containing the other six geographic regions, $\hat{\theta}$ is 0.290, indicating that the Arctic/Subarctic and the rest of North America constitute regions that are almost as genetically distinct as individual tribal groups. (This value of $\hat{\theta}$ was higher than that for three geographic regions: The Southwest/California/Great Basin, Arctic/Subarctic/Northwest Coast, and elsewhere.) This, of course, is primarily due to the virtual absence of haplogroups B, C, and D in the Arctic/Subarctic region, and the differences are even more pronounced when southern Athabaskans (Navaho and Apache) are eliminated from the estimate. Haplogroup A is the most common haplogroup in all of the geographic areas except the Southwest and California/Great Basin, where the frequencies of haplogroup A, 26% and 7%, respectively, are lower than in any other geographic area. When these three geographic groups, whose haplogroup distributions are most distinctive, are eliminated, the distributions of haplogroups A, B, and C + D among the remaining four geographic regions are much more similar but still statistically significantly heterogeneous ($\chi^2 = 29.7$, 6 d.f., $P < 0.001$).

The frequency of haplogroup B is highest in the Southwest region and declines, increasingly sharply, to the west, south, east, and, finally, to the north, where it is almost completely absent. The overall frequency of haplogroup B in North America reported in this study (32%) is greater than that reported in our earlier study (20% [Lorenz and Smith, 1994]). The exclusion of individuals who could not be assigned to one of the four

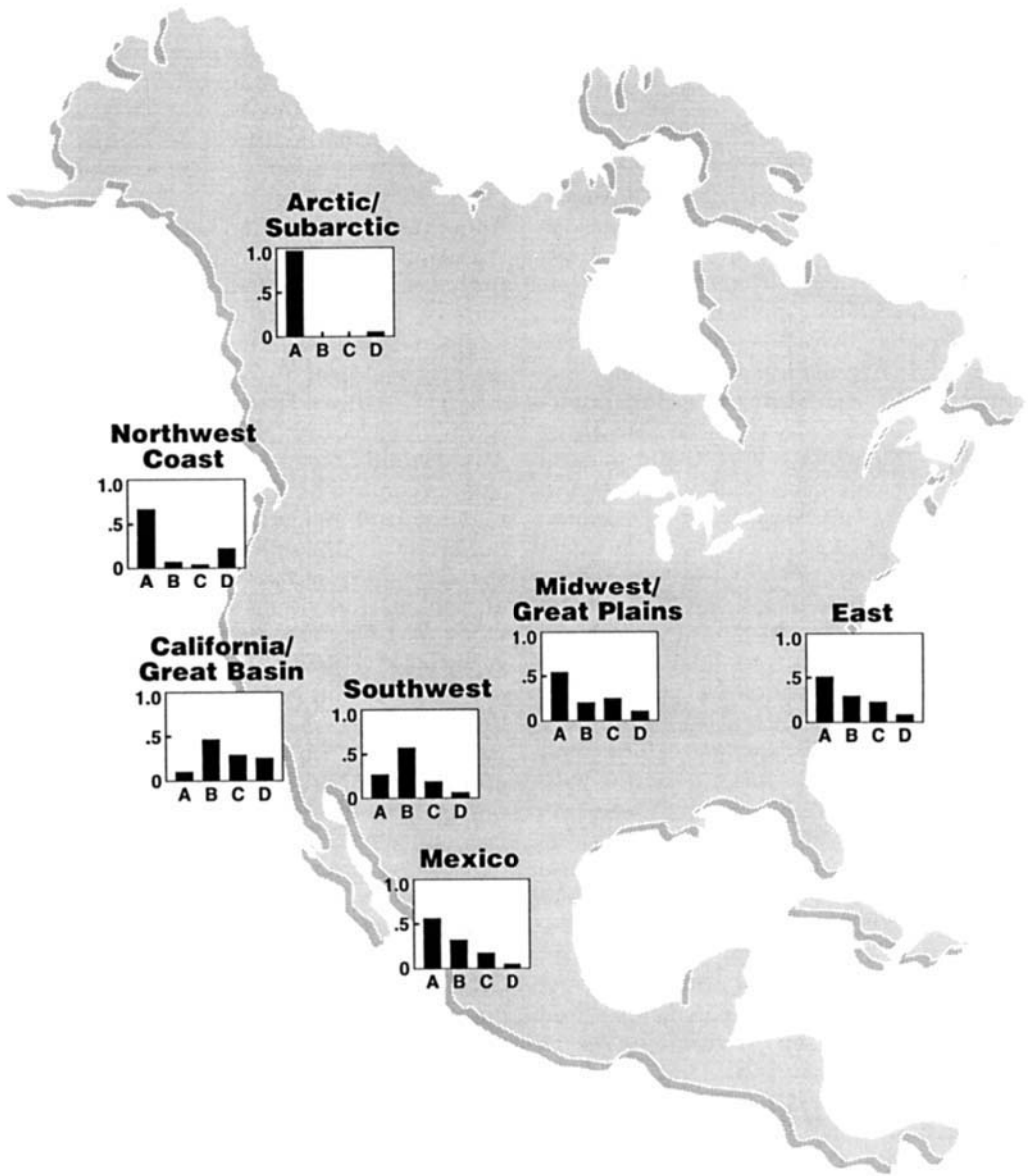


Fig. 2. Frequency distributions of the four mtDNA haplogroups by geographic region. The populations representing the geographic regions are as follows. Arctic/Subarctic: Inuit, Dogrib; Northwest Coast: Haida, Tlingit, Bella Coola, Nuu-Chah-Nulth; East: Cherokee, Choctaw, Chickasaw, Seminole, Creek, Narragansett, Micmac; Midwest/Great Plains: Chippewa, Kickapoo, Cheyenne, Arapahoe, Pawnee, Siouan, Kiowa, Coman-

che; Mexico: Maya, Paipai, Kiliwa, Cochimi, Nahua, Cora, Mixtec, Zapotec, Mixe; California/Great Basin: Hupa, Yurok, California Penutian, North California Hoka, Washo, Kumiai, Salinan, Chumash, Paiute, Shoshone, California Uto-Aztecan; Southwest: Navajo, Apache, Acoma, Zuni, Hualapai, Havasupai, Yavapai, Quechan, Mojave, Cocopa, Jemez, Taos, San Idelfonso, Pima, Hopi.

TABLE 3. Contribution of geography and language to clustering of haplogroup frequencies in North America

	$\hat{\theta}$ value ¹
Geography	
30 tribes	0.291
7 culture areas	0.193 (0.256) ²
3 geographic groups	0.199 (0.275) ²
2 geographic groups	0.290 (0.430) ²
Language	
24 languages	0.284
20 languages	0.242
9 language groups	0.205 (0.246) ²
3 language families	0.184 (0.322) ²
UPGMA clades	0.293 (0.322) ³

Source of homogeneity of UPGMA clades

Clade	Geography	Language
I	High	Low
II	Medium	Low
III	Low	Medium
IV	High	High

¹ Value of chi-square for homogeneity statistically significant at least the 0.05 level of probability.

² Value of $\hat{\theta}$ after eliminating Apache and Navaho from sample.

³ The five tribes identified by an asterisk in Figure 3 were excluded from this estimate.

haplogroups might have inflated the frequency of haplogroup B in this study. However, it is more likely that the difference is due to the fact that a larger number of individuals that were typed only for the presence or absence of the 9 bp deletion but not the other haplogroup markers in other studies (Shields et al., 1992, 1993; Ward et al., 1991, 1993) were necessarily excluded from the present but not from our earlier analysis. These individuals came from areas generally low in the frequency of haplogroup B.

Haplogroup C occurs at a frequency of 17% in North America and, while more homogeneously distributed than other haplogroups, is absent or nearly absent in the northwestern part of North America. Haplogroup D, the rarest of the four haplogroups in North America at 7%, is most prevalent in the northern Paiute/Shoshone, in which its frequency exceeds 40%. It is present at very low frequencies (<6.0%) everywhere except in the Northwest Coast and California/Great Basin regions, where it reaches frequencies of 18% and 22%, respectively. Even when the three most similar geographic regions (the East, Midwest/Great Plains, and Mexico) are compared, their haplogroup distributions are statistically significantly heterogeneous at the 0.05 level of probability ($\chi^2 = 10.8$, 4

d.f.). The only two of the seven geographic regions whose haplogroup frequencies are not statistically significantly different were Mexico and the East ($\chi^2 = 2.82$, 2 d.f., $P > 0.24$), underscoring the strong influence of geography on the distribution of haplogroups.

Greenberg (1987) divided all New World languages into three families: Eskimo-Aleut, Na-Dene, and Amerind. The distribution of the four haplogroups differs between the Eskimo-Aleut and northern Na-Dene groups (as represented by the Inuit, Haida, Tlingit, and Dogrib) on one hand and the Amerind groups and southern Na-Dene (Apache and Navajo) on the other ($\chi^2 = 137.5$, 3 d.f., $P < 0.001$). This results from the fact that the Eskimo-Aleut and northern Na-Dene individuals (but not, typically, those in Amerind groups) are almost exclusively haplogroup A. All of the 24 Amerind groups represented by more than ten individuals have haplogroups B and C (Table 2), as do the Navajo and Apache. Eight of the 24 Amerind groups lack haplogroup D, three lack haplogroup A, and four lack both haplogroups A and D.

$\hat{\theta}$ values for clusters of tribes based on a conservative linguistic taxonomy (Campbell and Mithun, 1979) (see Table 2) range from 0.284 (24 language groups), only slightly lower than that for individual tribal groups, to 0.242 (20 language groups). $\hat{\theta}$ values based on Greenberg's (1987) more inclusive linguistic taxonomy range from 0.205 (nine groups based on the following stocks: Eskimo, Haida, Continental Na-Dene, and the first sublevel groupings: Almosan-Keresiouan, Penutian, Hokan, Kiowa-Tanoan, Oto-Mangue, and Uto-Aztecan), a value slightly greater than that obtained for the seven geographic regions, to 0.184 for the three families (Eskimo-Aleut, Na-Dene, and Amerind). When the southern Athabaskan groups, who are presumed to share recent (i.e., as recently as 500 years ago) ancestry in common with northern Athabaskans (Basso, 1983; Brugge, 1983), are removed from the analysis, $\hat{\theta}$ values for Greenberg's (1987) nine language groups increases (from 0.205) to 0.246 ($\chi^2 = 337.9$, with 14 d.f., $P < 0.001$), and $\hat{\theta}$ for the three families increases (from 0.184) to 0.322 ($\chi^2 = 142.1$, with 4 d.f., $P < 0.001$).

The latter value is still lower than the comparable value (0.430) for only two geographic regions that combine all Na-Dene and all Eskimo-Aleut speakers into the Arctic/Subarctic group. While the haplogroup distribution of northern Na-Dene did not differ significantly from that of Eskimo-Aleuts, heterogeneity among the six haplogroup distributions (A, B, and C + D) of the Amerind language families alone is statistically significant ($\chi^2 = 156.0$, with 10 d.f., $P < 0.001$), and no two of the six haplogroup distributions are homogeneous at the 0.05 level of probability.

Genetic distance. The phenetic tree obtained by UPGMA divides the 30 modern North American tribes and the pre-Columbian Norris Farms population into four major clusters based on their frequencies of the four haplogroups. As illustrated in Figure 3, a major split in this tree is defined by the absence (or near absence) of haplogroup A in tribes within cluster IV. Thus, contrary to expectations based on the tripartite language taxonomy of Greenberg (1987), most Amerinds cluster together with Eskimo-Aleut and Athabascans in the other main branch. Cluster IV is also characterized by a high frequency of haplogroup B and includes both Hokan-speaking and Uto-Aztecan-speaking tribes, each occupying at least two of the seven geographic regions including or adjacent to the Southwest. Only one of three Uto-Aztecan-speaking and one of seven Hokan-speaking tribes falls outside cluster IV. Thus, cluster IV is characterized by relative homogeneity of both language and geographic origin. A curious exception is the presence of Cherokee in this cluster. The single Kiowa-Tanoan group, which Greenberg (1987) regards as closely related, together with Oto-Mangue-speaking tribes, to Uto-Aztecan, also falls in this cluster.

The remaining tribes are subdivided into three clusters. Cluster I comprises the three northwesternmost groups that are predominantly, if not exclusively, of haplogroup A. With the exception of southern Athabascans, this includes all members of two of the three major families of (presumably) unrelated languages (i.e., Eskimo-Aleut and Na-Dene). Thus, tribes in cluster I exhibit geographic

but not language homogeneity. Cluster II consists of tribes that have haplogroups A, B, and C but not D. Three of the four tribes from Mexico used in this analysis and both southwestern tribes not included in cluster IV fall in this cluster. As these six tribes that share close geographic proximity represent four different language groups, cluster II is primarily geographically defined. Cluster III consists of tribes that have all four haplogroups and was the most geographically heterogeneous of the four clusters, including tribes from six of the seven geographic regions. In contrast, six of the eight tribes in this cluster belonged to either the Penutian or Almosan-Keresiouan language group. Thus, language is more instrumental than geography in structuring this cluster. A maximum likelihood tree gives virtually the same four clusters as the UPGMA tree, although five groups (Mixtec, Zapotec, Chipewya/Kickapoo, Zuni, and California Penutians) are placed in different clusters. When these five tribal groups are excluded, the θ value associated with the UPGMA quadripartite phenetic clustering of tribal groups is 0.322, a value as high as that associated with any clustering based on language or geographic proximity except that for two geographic regions. None of the tribes in clusters I, III, or IV were placed in different clusters by the two methods.

DISCUSSION

A comparison of the distribution of the four haplogroups in North America with that in the other regions that have been studied is provided in Figure 4. Almost all Native North Americans (94%) belong to one of the four haplogroups, as do Native Central and South Americans (Torroni et al., 1992, 1993a, 1994b; Baillet et al., 1994). However, the frequencies of the haplogroups vary significantly but not clinally among North, Central, and South America ($\chi^2 = 257.45$, 6 d.f., $P < 0.001$), suggesting that the populations in North and South America have evolved relatively independently with restricted gene flow through Central America.

Greenberg's (1987) classification of all Native North American languages into three major families, and especially his hypothesis of "the unity of Amerind," has met with much

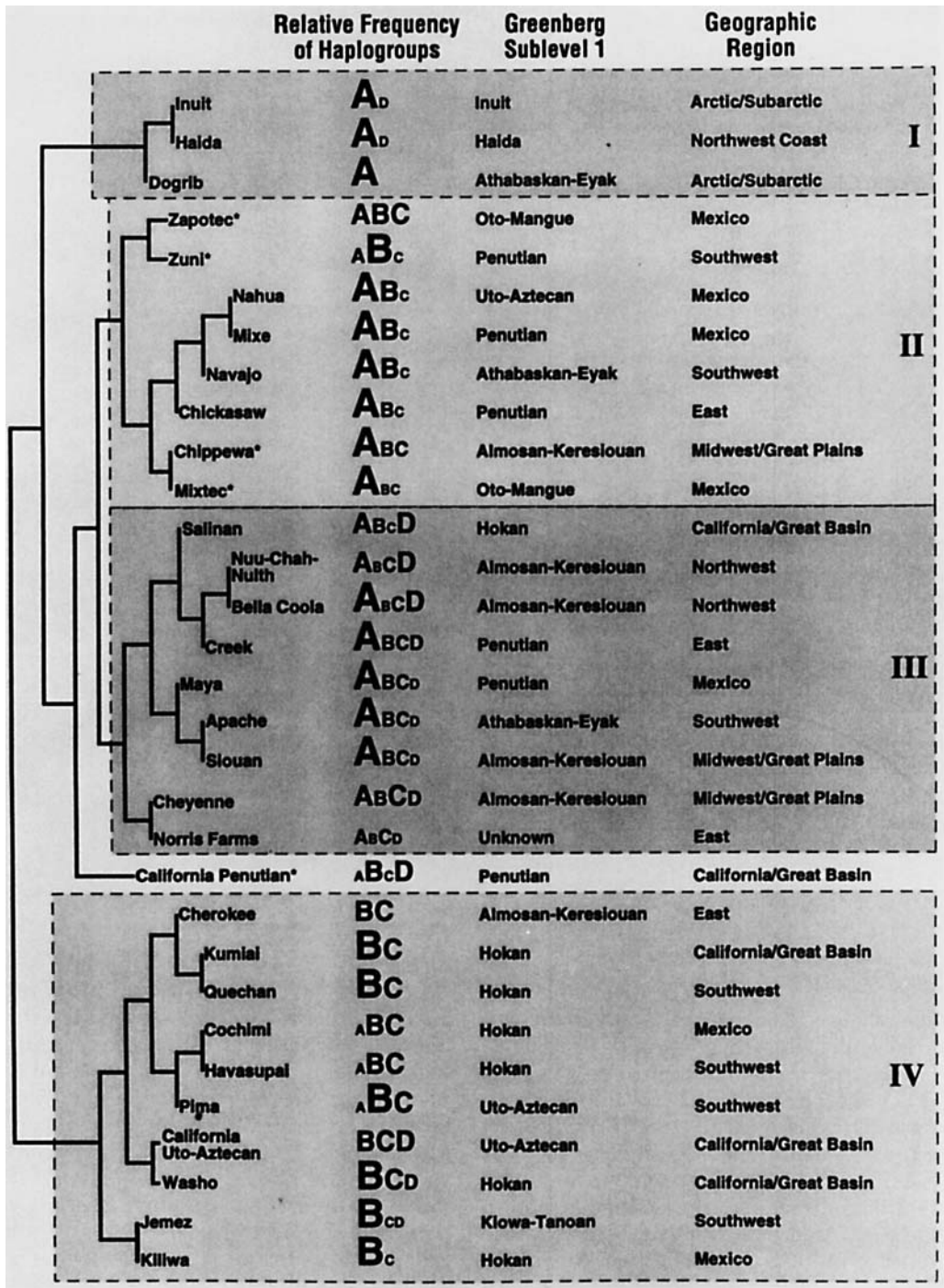


Fig. 3. Phenetic tree constructed in PHYLIP (Felsenstein, 1993) using a UPGMA algorithm from genetic distances calculated according to Cavalli-Sforza and Edwards (1967). The relative frequencies of the four haplogroups in each population are indicated by the size of the letter: largest letter, 51–100%; next largest, 26–

50%; next to smallest, 11–25%; smallest, 1–10%. An asterisk denotes populations that are found in different clusters in the maximum likelihood tree. Since this paper was submitted, haplogroup C has also been identified (Merriwether et al., 1995) in Dogrib people.

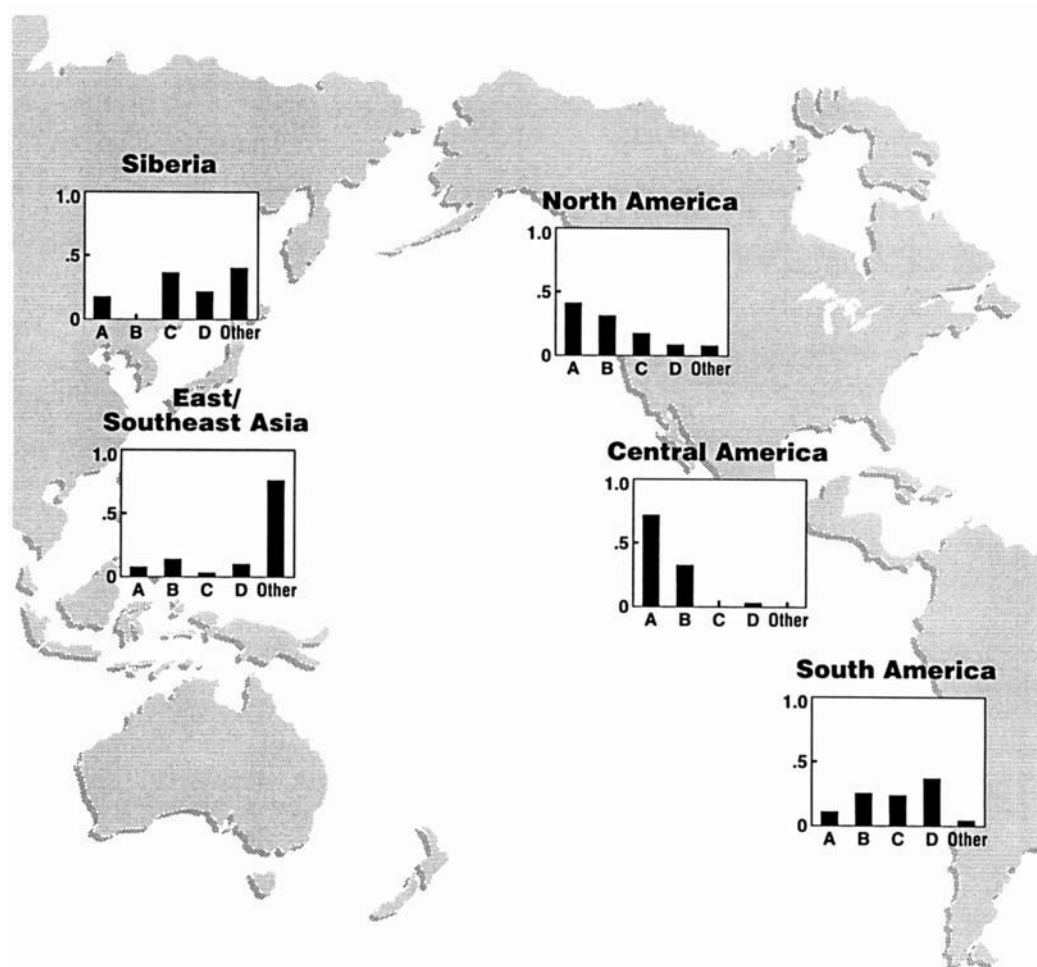


Fig. 4. Frequency distributions of mtDNA haplogroups in Asia and the New World.

resistance among comparative linguists. Nevertheless, some dental and genetic evidence supporting his tripartite division of Native Americans has been used to argue for three separate migrations into the New World (Greenberg et al., 1986). The present analysis, while not designed to test Greenberg's hypothesis, points to a fundamental split among the Eskimo/northern Na-Dene, Amerinds from the southwestern states of the U.S., and all other Amerinds. The clustering of the Eskimo and northern Na-Dene, who speak unrelated languages, together (cluster I) in the phenetic and $\hat{\theta}$ analyses is consistent with the conclusions of previous genetic studies (Szathmary and

Ossenberg, 1978; Shields et al., 1993; Szathmary, 1996) and does not support the three migration hypothesis. Although most Amerind groups, including those in Central and South America (Merriwether et al., 1995), fall in clusters II–IV, the major division in the phenetic tree separated most Hokan-speaking and Uto-Aztecan-speaking Amerinds (cluster IV), whose languages are not regarded as closely related, from all other groups. Moreover, no two Amerind language families exhibited homogeneous haplogroup distributions. Thus, our data do not support "the unity of Amerind" (Greenberg, 1987).

Except as noted above, $\hat{\theta}$ values decrease as the 30 tribes are lumped into more inclu-

sive linguistic taxa. However, the relative level of intergroup heterogeneity, as measured by the frequencies of the four mtDNA haplogroups, is about the same whether the tribes are grouped conservatively into 20 linguistic taxa or into Greenberg's (1987) nine major sublevels (when the southern Na-Dene are removed from the analysis). This indicates Greenberg's linguistic taxonomy produces clusters that are just as genetically homogeneous as when they are clustered by a more conservative linguistic taxonomy. In a study of sequence variation in the mtDNA control region, Ward et al. (1993) found a lack of congruence between the linguistic taxonomy of three northwest coast tribes, the Haida, Bella Coola, and Nuu-Chah-Nulth, and the distribution of their mtDNA types (although the Haida exhibited less intratribal variation), suggesting a strong influence of geography but not language on genetic similarities in that region. The geographic contribution to genetic similarities in the present study is illustrated in Figure 1. Here, the geographic homogeneity among members of clusters I (triangles in the northwest) and IV (open circles in the southwest) and the Mexican tribes included in cluster II (squares in southern Mexico) is well defined. However, all four clusters in the phenogram of tribes in the present study (illustrated in Fig. 3) include tribes of different language and geographic origins. Although the present study shows that there is no clear clustering among most Amerind language taxa or geographic groups, there are some instances where either language or geography is more highly correlated with genetic differences.

The clustering of adjacent tribes whose languages are unrelated in clusters I and IV probably results from admixture. In contrast, the tribes in cluster III are more geographically diverse than those in any of the other three clusters. Language spreads among genetically unrelated (or only remotely related) tribes might have been more influential in shaping linguistic diversity among tribes in cluster III than elsewhere. In contrast, the persistence of language in other regions, despite close intergroup contact, is reflected in the fact that the two southern Athabaskan tribes studied fall in

separate clusters, neither of which is identical to that including northern Athabaskans. The geographic clustering of tribes of Mexico, representing three different language groups, in cluster II might reflect the greater importance of gene flow in shaping the haplogroup distribution in Mexico than in Amerind tribes to the north. All groups classified by Greenberg (1987) as Hokan, with the exception of the Salinan/Chumash of California, fall in cluster IV, suggesting restricted gene flow with other language groups. However, with the exception of the Cherokee, all tribes in cluster IV are located in the Southwest and surrounding areas (southern California/Great Basin and Baja California) reflecting a (less pronounced) geographic component to the distribution of the mtDNA haplogroups (Fig. 3). Paralleling the clustering of most Hokan-speaking groups with Uto-Aztecan-speaking groups in close proximity to them is their near exclusive sharing of Al^{Me}, a rare mutation at the albumin locus (Schell and Blumberg, 1988; Smith et al., in preparation). The relative contributions of gene flow and common ancestry to this close genetic relationship are currently under investigation using DNA sequencing analysis (Lorenz and Smith, in preparation) and microsatellite (SSR) (see Morin and Smith, 1995) polymorphisms linked to the albumin locus (Smith et al., in preparation).

The present study concurs with that of Torroni et al. (1994b), which also found that the Maya and the Mixe, included by Greenberg (1987) in the Penutian sublevel, did not cluster together genetically. On the other hand, Torroni and Wallace (1994) and Torroni et al. (1994b) have argued that the mtDNA of Uto-Aztecs (represented by Pimas) and Oto-mangueans (represented by Zapotecs and Mixtecs) clusters together and thus lends genetic support to Greenberg's (1987) hypothesis that the two (along with Kiowa-Tanoan) comprise the Central Amerind Stock. However, the present study shows that when data from additional Uto-Aztecan and Tanoan groups are included, genetic support for Greenberg's Central Amerind language taxonomy is not as strong. Figure 3 shows that although California Uto-Aztecs (as well as Kiowa-Tanoans, also

included by Greenberg in the Central Amerind language stock) do cluster with the Pima from the Southwest in group IV, the Nahua, a Uto-Aztecan-speaking group from Mexico, as well as the Oto-manguean groups, Zapotec and Mixtec, cluster with other Mexican (as well as non-Mexican) groups in group II.

The Southwest area is interesting because the tribes in this region have the highest frequencies of haplogroup B in North America and are virtually lacking in haplogroup A, the most common haplogroup in North America. The groups in the Southwest with the highest frequencies of haplogroup A, which is nearly fixed in northern Na-Dene groups, are, not surprisingly, the southern representatives of Na-Dene, the Navajo and Apache. A recent study of a prehistoric southwestern population (Parr et al., 1996) suggests that haplogroups A and B were absent and very common, respectively, in the US Southwest before the Navaho and Apache are believed to have arrived there. The Navajo and Apache also exhibit relatively high frequencies of haplogroups B and C which are very rare (though not completely absent [Merriwether et al., 1995]) in northern Na-Dene groups and, barring the loss of both haplogroups due to a severe bottleneck within the last millennium, were also very rare in the ancestors of the Na-Dene. Ethnohistoric records document extensive trading, raiding, and emigration among the Apache, Navajo, and other southwestern groups (Basso, 1983; Brugge, 1983). The distribution of mtDNA haplogroups in the Southwest suggests that this contact was accompanied by extensive female migration among the groups. Let us assume that the presence of haplogroups B and C in the Navajo and Apache is completely due to gene flow from the groups surrounding them in the Southwest and that the presence of haplogroup A in the southwestern groups is completely due to gene flow from the Navajo and Apache. The former assumption is based on the near absence of haplogroups B and C in northern Na-Dene and the recency of the split of Navaho/Apache from Northern Na-Dene (Basso, 1983; Brugge, 1983); the latter assumption is justified by the apparent absence of haplogroup A in the Southwest before but not after the arrival of southern

Athabaskans (Parr et al., 1996). The rate of female gene flow (m) per generation required to produce the frequencies of haplogroups B and C observed in the southern Na-Dene and of haplogroup A observed in non-Na-Dene in the Southwest can be calculated by solving for m given that $(1 - m)^n = q_n - Q / q_0 - Q$. Here, n is the number of generations, q_n is the frequency of the allele in the present-day recipient population, q_0 is the frequency of the allele in the ancestral recipient population, and Q is the frequency of the allele in the donor population (Workman, 1973). Assuming the Apache and Navajo have been in the Southwest for approximately 500 years (Gunnerson, 1979) and 20 years/generation (i.e., $n = 25$), approximately 2.1% of the female breeding population each generation originated in the surrounding southwestern groups (for this calculation, based on the combined frequency of haplogroups B and C, $q_n = 0.412$, $q_0 = 0$, and $Q = 1$). In contrast, assuming equal population size of both groups, only 0.2% of non-Na-Dene southwestern females per generation were derived from the southern Athabaskan tribes (based on the frequency of haplogroup A where $q_n = 0.052$, $q_0 = 0$, and $Q = 1$). There has apparently been more female gene flow from the surrounding southwestern groups into the Apache and Navajo than from the Apache and Navajo into surrounding southwestern groups. Although either a small Apache/Navajo population (relative to that from which they acquired the B and C haplogroups) or a nonsouthwestern source for the B and C haplogroups (e.g., retention of primitive traits now lost in northern Na-Dene) rather than asymmetric mtDNA gene flow could be responsible, the latter explanation is in agreement with ethnohistorical records of gene flow of emigrants and captives from other southwestern groups into Navaho (Brugge, 1983) and Apache (Basso, 1983) groups. The hypothesis of asymmetric gene flow is also consistent with data on the distribution of albumin polymorphisms in the Southwest. The Apache and Navajo have both the Naskapi (Al^{Na}) and Mexico (Al^{Me}) albumin variants, while most Pima and Yuman groups in the Southwest (except in very rare instances) have only Al^{Me} (Schell and

Blumberg, 1988; Smith et al., in preparation). Since Al^{Na} originated in northern groups and was first introduced in the Southwest by southern Athabaskans and Al^{Me} originated in indigenous southwestern groups, the direction of gene flow must have been principally as hypothesized above. Moreover, since north of Mexico Al^{Me} is otherwise restricted to Uto-Aztecans and/or Yuman-speaking groups, they are likely sources for this admixture. That Al^{Na} is present, albeit very rare, in some Uto-Aztecanspeaking (e.g., Hopi) and Yuman-speaking (e.g., Maricopa and Mohave) tribes but in much lower frequencies than Al^{Me} reaches in southern Athabaskans is also consistent with our hypothesis of asymmetric gene flow.

The presence of Cherokee in cluster IV seems enigmatic especially since the Sioux, to whom they are regarded as closely related based on comparative linguistic data (Campbell and Mithun, 1979), fall in group III with most Algonquian-speaking groups to whom Greenberg (1987) hypothesizes they are closely related. This difference is primarily due to the high frequency of haplogroup A (which is nearly fixed in Athabaskans and common in Algonquians) in Siouan tribes and its absence in the Cherokee. Al^{Na} , a marker found in most Athabaskan- and Algonquian-speaking tribes (Schell and Blumberg, 1988; Smith et al., in preparation), is also often present, albeit in low frequencies, in Siouan but absent in Cherokee (as well as Mohawk, also speakers of an Iroquoian language) tribes, suggesting considerable admixture between Siouan but not Iroquoian tribes and Algonquian-speaking tribes. Athabaskans are unlikely sources of Al^{Na} in Siouan groups, assuming admixture was reciprocal, since they completely lack haplogroup B which is common to both Siouan- and Algonquian-speaking groups. Such admixture must have postdated the divergence between Siouan-speaking and Iroquoian-speaking (e.g., Cherokee and Mohawk) peoples and the migration of the latter eastward and could be associated with the migration of Algonquian speakers from the Columbia Plateau eastward through the Plains (Denny, 1991) or their subsequent expansion through, and out of, southern On-

tario (Siebert, 1967), which probably occurred within the last 3,000 years.

When data from the few archaeological populations in North America that have been studied for mtDNA variation are compared with that provided by the present study, continuity between ancient populations in an area and present-day tribes can be assessed. Stone and Stoneking (1993) found that 49 out of 50 individuals from a cemetery at Norris Farms in central Illinois dating to AD 1300 belonged to one of the four haplogroups. The Norris Farms population clusters in group III with other tribes from the East and Midwest/Great Plains. Interestingly, a Salt Lake Fremont population from Utah studied by Parr et al. (1996) lacks haplogroup A and exhibits a high frequency of haplogroup B which would place it in group IV where most present-day tribes in that region cluster. The distribution of haplogroups in prehistoric peoples in the western Great Basin do not differ from that of northern Uto-Aztecanspeaking peoples currently living in this region (Kaestle, 1995). These archaeological populations from three different regions suggest a long-term stability in the distribution of the four haplogroups, hence their utility in assessing ancestor-descendant relationships, in at least some regions of North America. The data provided in the present study might also make it possible to determine certain instances where lack of such stability suggests major population replacements.

CONCLUSIONS

The results of this study indicate that mtDNA haplogroup frequencies are not homogeneous among Native American tribes. Similarity in the frequency of the four mtDNA haplogroups among North American tribes is due both to shared ancestry as indicated by language affiliation (e.g., among Hokan-speaking groups) as well as gene flow between geographically proximate groups (e.g., between Siouan- and Algonquian-speaking groups). In some cases, such as that of gene flow between Yuman and/or Uto-Aztecanspeaking tribes and the southern Athabaskans, the relative contributions of shared ancestry and geographic proximity can be estimated. In other cases, such as that

of the similarity between Hokan and some Uto-Aztecan-speaking groups, hypotheses about language spreads or genetic relationships between languages based on comparative linguistic, archaeological, or more detailed genetic analysis are plausible. In still other cases, groups with demonstrably unrelated languages (Eskimo-Aleut and Na-Dene) share similar mtDNAs. This study greatly expands our knowledge of the frequencies of the four founding mtDNA haplogroups in North America and thus provides both an improved database for testing hypotheses regarding the overlap of linguistically defined cultural units and genetically defined groups and a baseline with which to compare samples collected from archaeological populations.

ACKNOWLEDGMENTS

This study was supported by grants RR00169, HD-8-2909, and RR05090 from the National Institutes of Health. We are indebted to numerous personnel of Indian Health Service Facilities and to all the people who have provided samples used in this study. Without their cooperation this study would not be possible. We also thank Dr. John Johnson (Santa Barbara Museum of Natural History) and Mr. David L. Schmidt (University College of Cape Breton, Nova Scotia) for providing hair samples.

LITERATURE CITED

- Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, and Young IG (1981) Sequence and organization of the human mitochondrial genome. *Nature* 290:457-465.
- Bailliet G, Rothhammer F, Carnese FR, Bravi CM, and Bianchi NO (1994) Founder mitochondrial haplotypes in Amerindian populations. *Am. J. Hum. Genet.* 54:27-33.
- Ballinger SW, Schurr TG, Torroni A, Gan YY, Hodge JA, Hassan K, Chen K-H, and Wallace DC (1992) Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient Mongoloid migrations. *Genetics* 130:139-152.
- Basso KH (1983) Western Apache. In WC Sturtevant (ed.): *Handbook of North American Indians: Volume 10 Southwest*. Washington, DC: Smithsonian Institution, pp. 462-488.
- Brugge DM (1983) Navaho prehistory and history to 1850. In WC Sturtevant (ed.): *Handbook of North American Indians: Volume 10 Southwest*. Washington, DC: Smithsonian Institution, pp. 489-501.
- Campbell L, and Mithun M (1979) North American Indian historical linguistics in current perspective. In L Campbell and M Mithun (eds.): *The Languages of Native America: A Historical and Comparative Assessment*. Austin: University of Texas Press, pp. 3-69.
- Cavalli-Sforza LL, and Edwards AW (1967) Phylogenetic analysis: Models and estimation procedures. *Evolution* 21:550-570.
- Chakraborty R, and Weiss K (1991) Genetic variation of the mitochondrial DNA genome in American Indians is at mutation-drift equilibrium. *Am. J. Phys. Anthropol.* 86:497-506.
- Denny JP (1991) The Algonquian Migration From Plateau to Midwest: Linguistics and Archaeology. Papers of the 22nd Algonquian Conference. Ottawa: Carleton University.
- Felsenstein J (1993) PHYLIP (Phylogeny Inference Package), version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle.
- Greenberg JH (1987) *Language in the Americas*. Stanford: Stanford University Press.
- Greenberg JH, Turner CG II, and Zegura SL (1986) The settlement of the Americas: A comparison of the linguistic, dental, and genetic evidence. *Curr. Anthropol.* 27:477-497.
- Gunnerson JH (1979) Southern Athapaskan Archeology. In WC Sturtevant (ed.): *Handbook of North American Indians: Volume 9 Southwest*. Washington, DC: Smithsonian Institution, pp. 162-169.
- Johnston FE, Blumberg BS, Agarwall SS, Melartin L, and Burch TA (1969) Alloalbuminemia in Southwestern U.S. Indians: Polymorphism of albumin Naskapi and albumin Mexico. *Hum. Biol.* 41:263-270.
- Kaestle F (1995) Mitochondrial DNA evidence for the identity of the descendants of the prehistoric Stillwater Marsh population. In CS Larsen and RL Kelly (eds.): *Bioarchaeology of the Stillwater Marsh: Prehistoric Human Adaptation in the Western Great Basin*. Anthropological Papers of the American Museum of Natural History, pp. 73-80.
- Lorenz JG, and Smith DG (1994) Distribution of the 9 bp mitochondrial DNA region V deletion among North American Indians. *Hum. Biol.* 66:777-788.
- Lorenz JG, and Smith DG (in preparation) Geographic and linguistic distribution of sequence variation in the mtDNA control region of native North American Indians.
- Merriwether DA, Rothhammer F, and Ferrell RE (1995) Distribution of the four founding lineage haplotypes in Native Americans suggests a single wave of migration for the New World. *Am. J. Phys. Anthropol.* 98:411-430.
- Morin PA, and Smith DG (1995) Nonradioactive detection of hypervariable simple sequence repeats in short polyacrylamide gels. *Biotechniques* 19:223-227.
- Nichols J (1992) *Linguistic Diversity in Space and Time*. Chicago: The University of Chicago Press.
- Parr RL, Carlyle SW, and O'Rourke DH (1996) Ancient DNA analysis of Fremont Amerindians of the Great Salt Lake wetlands. *Am. J. Phys. Anthropol.* 99:507-518.
- Schell LM, and Blumberg BS (1988) Alloalbuminemia and the migrations of Native Americans. *Yrbk. Phys. Anthropol.* 31:1-13.

- Schurr TG, Ballinger SW, Gan Y-Y, Hodge JA, Merriwether DA, Lawrence DN, Knowler WC, Weiss KM, and Wallace DC (1990) Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies, suggesting they derived from four primary maternal lineages. *Am. J. Hum. Genet.* 46:613-623.
- Shields GF, Hecker K, Voevoda MI, and Reed JK (1992) Absence of the Asian-specific region V mitochondrial marker in native Beringians. *Am. J. Hum. Genet.* 50:758-765.
- Shields GF, Schmiechen AM, Frazier BL, Redd A, Voevoda MI, Reed JK, and Ward RH (1993) mtDNA sequences suggest a recent evolutionary divergence for Beringian and northern North American populations. *Am. J. Hum. Genet.* 53:549-562.
- Siebert FT Jr (1967) The original home of proto-Algonkian people. *Anthropological Series* 78. *Natl. Mus. Can. Bull.* 214:13-47.
- Siegel S (1956) *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw Hill Book Co., Inc.
- Smith DG, Lorenz JG, Kanthaswamy S, Rolfs BK, and Bettinger RL (in preparation) Ethnohistorical implications of the distributions of albumin Naskapi and albumin Mexico in North America.
- Stone AC, and Stoneking M (1993) Ancient DNA From a pre-Columbian Amerindian population. *Am. J. Phys. Anthropol.* 92:463-471.
- Szathmary EJE (1996) Ancient migrations from Asia to North America. In T Akazawa and EJE Szathmary (eds.): *Prehistoric Mongoloid Dispersals*. Oxford: Oxford University Press, pp. 149-164.
- Szathmary EJE, and Ossenberrg NS (1978) Are the biological differences between North American Indians and Eskimos truly profound? *Curr. Anthropol.* 19:673-685.
- Torroni A, and Wallace DC (1994) Mitochondrial DNA variation in human populations and implications for detection of mitochondrial DNA mutations of pathological significance. *J. Bioenerg. Biomembr.* 26: 261-271.
- Torroni A, and Wallace DC (1995) Mitochondrial DNA haplogroups in Native Americans. *Am. J. Hum. Genet.* 56:1234-1236.
- Torroni A, Schurr TG, Yang C-C, Szathmary EJ, Williams RC, Schanfield MS, Troup GA, Knowler WC, Lawrence DN, Weiss KM, and Wallace DC (1992) Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. *Genetics* 130:153-162.
- Torroni A, Schurr TG, Cabell MF, Brown MD, Neel JV, Larsen M, Smith DG, Vullo CM, and Wallace DC (1993a) Asian affinities and continental radiation of the four founding Native American mtDNAs. *Am. J. Hum. Genet.* 53:563-590.
- Torroni A, Sukernik RI, Schurr TG, Starikovskaya YB, Cabell MF, Crawford MH, Comuzzie AG, and Wallace DC (1993b) mtDNA variation of aboriginal Siberians reveals distinct genetic affinities with Native Americans. *Am. J. Hum. Genet.* 53:591-608.
- Torroni A, Miller JA, Moore LG, Zamudio S, Zhuang J, Droma T, and Wallace DC (1994a) Mitochondrial DNA analysis in Tibet. Implications for the origin of the Tibetan population and its adaptation to high altitude. *Am. J. Phys. Anthropol.* 93:189-199.
- Torroni A, Chen Y-S, Semino O, Santachiara-Beneceretti AS, Scott CR, Lott MT, Winter M, and Wallace DC (1994b) mtDNA and y-chromosome polymorphisms in four Native American populations from southern Mexico. *Am. J. Hum. Genet.* 54:303-318.
- Wallace DC, Garrison K, and Knowler WC (1985) Dramatic founder effects in Amerindian mitochondrial DNAs. *Am. J. Phys. Anthropol.* 68:149-155.
- Ward RH, Frazier BS, Dew K, and Paabo S (1991) Extensive mitochondrial diversity within a single Amerindian tribe. *Proc. Natl. Acad. Sci. U.S.A.* 88:8720-8724.
- Ward RH, Redd A, Valencia D, Frazier B, and Paabo S (1993) Genetic and linguistic differentiation in the Americas. *Proc. Natl. Acad. Sci. U.S.A.* 90:10663-10667.
- Weir BS (1990) Intraspecific differentiation. In DM Hillis and C Moritz (eds.): *Molecular Systematics*. Sunderland, Massachusetts: Sinauer Assoc., Inc., pp. 373-410.
- Workman PL (1973) Genetic analysis of hybrid populations. In MH Crawford and PL Workman (eds.): *Methods and Theories of Anthropological Genetics*. Albuquerque: University of New Mexico Press, pp. 117-150.